

# Ferramentas de informática para análise de resultados

Maria Margarida Ribeiro

Unidade Departamental de Silvicultura e Recursos Naturais

Escola Superior de Castelo Branco, 6000 Castelo Branco. Email: [mribeiro@esa.ipcb.pt](mailto:mribeiro@esa.ipcb.pt)

## Sumário

Um progresso rápido tem sido feito em muitas áreas da biologia molecular devido à utilização dos microsátélites, acompanhado por uma revolução ao nível da sua análise estatística. Esta apresentação revê algumas das propriedades dos microsátélites que os tornam tão apetecíveis, descreve alguns modelos teóricos de mutação importantes para a dedução de estatísticas mais precisas, nomeia alguns problemas devidos à utilização dos microsátélites, exemplifica quais as principais aplicações destes marcadores e sumariza alguns dos pacotes estatísticos mais recentes que podem ser utilizados para analisar dados gerados por microsátélites.

Palavras-chave: microsátélites, pacotes estatísticos, modelos de mutação, biologia molecular

## Summary

A fast progress is occurring in molecular biology as a result of the widespread use of microsatellites, followed by a parallel statistical analysis revolution. The foregoing paper reviews some microsatellites properties that have made them preferred markers, describes some theoretical mutation models fundamental for improving statistics' accuracy, refers some problems related with microsatellite use, exemplifies the main application of those markers and summarises some recent statistical packages for analysing microsatellite-based data.

Keywords: microsatellites, statistical packages, mutation models, molecular biology

## Introdução

Os investigadores da área da genética molecular tentam utilizar marcadores moleculares muito variáveis e que produzam resultados com facilidade. Esta atitude é explicada pela necessidade de extrair o máximo de informação possível de cada experiência, que envolve normalmente uma quantidade enorme de amostras e imenso trabalho no laboratório. Após a obtenção dos resultados laboratoriais, começa a questão do significado desse amontoado de dados. Para que os dados obtidos possam ser analisados em profundidade, temos que ter presente as questões que levaram ao estudo e que tipo de inferências pretendemos extrair: heterozigocidade, sistema de cruzamento, estrutura da espécie, filogenia, etc.

Existe uma revolução em curso ao nível da ecologia molecular, da genética de populações e da conservação dos recursos genéticos, devido à utilização em larga escala dos microsátélites (SSRs – “simple sequence repeats”). Os microsátélites têm sido preferidos relativamente a outros marcadores moleculares devido ao seu elevado polimorfismo, ocorrência frequente e distribuição uniforme ao longo do genoma, leitura fácil e precisa, herança codominante e possibilidade de automatização. No entanto, muitos investigadores ainda não se aperceberam da revolução paralela que tem vindo a acontecer ao nível da análise estatística, como resposta à quantidade e qualidade dos dados gerados pelos SSRs (Luikart & England 1999).

Nesta apresentação vou considerar os seguintes aspectos: a) rever a natureza dos marcadores que podemos utilizar para avaliar a variação genética, b) considerar algumas características dos microsátélites relevantes para o estudo da variação genética e também algumas das suas limitações, c) esquematizar a organização do trabalho de modo a que as questões postas sejam respondidas e fazer

referência a alguns dos programas recentes usados para analisar dados obtidos com SSR, d) realçar, com a ajuda de um exemplo, o encaixe das perguntas com os pacotes estatísticos utilizados e as respostas obtidas.

## Marcadores

O conhecimento da estrutura genética de uma espécie pode ser obtido com base em marcadores polimórficos que permitem a determinação das frequências génicas e genótípicas. A variação genética das espécies também pode ser estudada usando a análise quantitativa das características morfológicas, mas devido à influência ambiental, ao carácter poligénico de algumas características e ao tempo e custo necessário para obter a informação, outros métodos têm vindo a ser desenvolvidos (Wang & Szmidt 2001).

Um *marcador molecular* pode ser definido como uma sequência de ADN ou uma proteína que pode ser detectada e cuja heritabilidade pode ser inferida. É o polimorfismo dos marcadores moleculares que pode ser usado para estudar a diversidade genética. O polimorfismo das proteínas tem sido estudado através das alozimas, i.é., diferentes formas moleculares de uma enzima codificadas por diferentes alelos de um *locus*. O polimorfismo pode, também, ser identificado em diferentes tipos de ADN: nuclear ou citoplásmico (no cloroplasto e na mitocôndria) (Mitton 1994; Parker *et al.* 1998; Vekemans & Jacquemart 1997)). Vários tipos de marcadores podem ser usados: marcadores codominantes, alozimas, RFLPs (Restriction Fragment Length Polymorphisms), SNPs (Single Nucleotide Polymorphisms), SSRs (Simple Sequence Repeats) e marcadores dominantes, RAPDs (Random Amplified Polymorphic DNA) e AFLPs (Amplification Fragment Length Polymorphism), entre outros.

As propriedades desejáveis de um marcador são: a codominância (as diferentes formas de um marcador devem ser detectáveis em organismos diplóides de modo a distinguir os homozigóticos dos heterozigóticos), a distribuição uniforme ao longo do genoma, a detecção fácil, rápida e pouco dispendiosa e resultados reprodutíveis dentro e entre laboratórios (aspectos revistos por Karp & Edwards (1997); Parker *et al.* (1998); Szmidt & Wang (2000)).

## Microsatélites

### *Características:*

Um novo tipo de marcador conhecido como *microsatélite* foi desenvolvido com base na variação da sequência do ADN. Este marcador é baseado em repetições tandem de motivos pequenos (de 1 a 6 pares de bases), repetido várias vezes (Tautz 1989). As diferenças no comprimento de um alelo de um *locus* SSR são detectadas por amplificação do DNA através da reacção em cadeia da polimerase (PCR), usando dois “primers” que complementem a sequência flanqueadora única do *locus*. O polimorfismo é detectado pela separação dos fragmentos por electroforese e o seu tamanho pode diferir somente num par de bases. Sugere-se que a variação no tamanho dos alelos é devida ao ‘escorregamento’ da ADN polimerase durante a replicação dos motivos, seguida da não reparação do AND desalinhado (Strand *et al.* 1993).

O tamanho do motivo parece influenciar a taxa de mutação e, por isso, motivos mais extensos são mais polimórficos que os mais curtos (Ellegren 2000; Primmer *et al.* 1998). Isto é provavelmente devido à maior estabilidade de configurações desalinhadas nas sequências de motivos mais longos. Por outro lado, microsatélites que contenham motivos interrompidos (devido à inserção ou substituição de bases) parecem ter taxas de mutação inferiores às de microsatélites perfeitos, o que parece ser devido a uma menor probabilidade de se produzirem formas desalinhadas quando surgem motivos interrompidos (Petes *et al.* 1997). A maioria dos microsatélites contém menos de algumas dezenas de unidades. Isto faz suspeitar a existência de fortes limitações ao crescimento da série de repetições: séries mais longas tornam-se mais instáveis (Ellegren 2000).

### *Modelos de mutação:*

O desenvolvimento de estatísticas que possam refletir de forma precisa, por exemplo, a estrutura genética de uma espécie, requer uma compreensão do modelo de mutação envolvido no processo de evolução dos SSRs (Balloux & Lugon-Moulin 2002). No modelo de infinitos alelos (‘infinite allele

model' - IAM, (Kimura & Crow 1964)) uma mutação pode envolver qualquer número de unidades repetidas e resulta sempre num novo estado alélico não existente previamente na população. Como consequência, este modelo não permite a homoplasia (isto é, dois alelos são iguais mas provêm de duas linhagens diferentes). O modelo passo a passo ('stepwise mutation model' - SMM, (Kimura & Ohta 1978)) descreve melhor o processo evolucionário destes marcadores. No SMM, cada mutação cria um novo alelo por adição ou deleção de uma unidade repetida, com igual probabilidade em ambas as direcções. Por consequência, alelos mais afastados em termos de tamanho também estão mais afastados do ponto de vista evolucionário do que alelos com tamanhos semelhantes: o SMM contém uma memória evolucionária. O modelo de K alelos (KAM) também pode ser considerado para os microsatélites. Neste caso, existem K possíveis formas alélicas e cada alelo tem a mesma probabilidade de mutar para cada um dos K-1 alelos (Crow & Kimura 1970). Devido à restrição no número de unidades repetidas possíveis dos alelos de um *locus* SSR, o KAM parece ser mais realístico do que o IAM.

Diferentes tipos de estimadores, baseados no SMM, têm sido desenvolvidos para estimar relações filogenéticas, distâncias genéticas e diferenciação das populações (Balloux & Lugon-Moulin 2002; Goldstein *et al.* 1995; Shriver *et al.* 1995; Slatkin 1995). Esses estimadores têm como base os seguintes pressupostos: i) as mutações resultam na alteração de um só motivo; ii) a taxa de mutação é constante e independente do tamanho da unidade repetida; iii) não existem limitações devidas ao tamanho do alelo. No entanto, têm sido observadas discrepâncias significativas entre tempos de divergência conhecidos e distâncias genéticas baseadas em SSR (Garza *et al.* 1995) o que significa que pelo menos um desses pressupostos está errado. Além disso, as taxas de mutação podem variar com o tamanho da unidade repetida (di-, tri- e tetranucleótidos), com a composição em bases da unidade repetida (Bachtrog *et al.* 2000), o tipo de microsatélite (perfeito, composto ou interrompido) e com o grupo taxonómico. Podem também variar com a natureza da região flangeadora, a posição do microsatélite no cromossoma e o tamanho dos alelos (os alelos maiores são mais susceptíveis de sofrerem mutações que os mais curtos) (Schlötterer *et al.* 1998).

### *Problemas:*

Os microsatélites têm alguns pontos fracos. Pode haver perda ou ausência de amplificação de alguns alelos devido à substituição de bases ou "indels" na região flangeadora, o que dá origem a alelos nulos. Um heterozigótico com um alelo nulo não pode ser distinguido, no gel, do homozigótico que possui o alelo que foi amplificado. Isto tem como consequência uma subestimação da heterozigocidade em relação ao que se poderia esperar em condições de equilíbrio de Hardy-Weinberg. Outro problema está associado com o escorregamento da *Taq* polimerase durante a PCR e gera problemas na leitura das bandas em particular se se utiliza um método automatizado (Liepelt *et al.* 2001). Existe, também, o fenómeno de 'drop-out', que ocorre quando a quantidade de DNA é insuficiente para uma boa amplificação e, como consequência, só é amplificado o mais curto dos dois alelos (Taberlet & Luikart 1999): o que implica, também, uma subestimação da homozigocidade. A hipervariabilidade dos marcadores nem sempre representa uma vantagem. Fragmentos de tamanho idêntico podem não ser idênticos na sua sequência ancestral, o que possibilita a homoplasia, como já foi referido. Este fenómeno torna menos precisas estimativas de parâmetros de genética de populações (assunto revisto por Estoup *et al.* (1995)). A homoplasia aumenta com a taxa de mutação e o tempo de divergência. Verifica-se um aumento na heterozigocidade dentro das populações e uma subestimação da diferenciação genética entre populações. Além disso, reduções no efectivo populacional podem conduzir a distâncias genéticas muito grandes num curto espaço de tempo (Hedrick 1999). O mesmo autor sugere que é necessário avaliar os dados obtidos com *loci* muito polimórficos - SSR -, porque a informação que eles nos fornecem pode ser bastante diferente da obtida com *loci* menos polimórficos, como se exemplifica em Balloux *et al.* (2000). No entanto, no artigo de Estoup *et al.* (1995) conclui-se que a homoplasia só se torna verdadeiramente problemática quando envolve elevadas taxas de mutação, um efectivo populacional muito grande e limitações fortes no tamanho dos alelos. No entanto, a homoplasia tem sido considerada em várias distâncias genéticas que são baseadas no SMM (Feldman *et al.* 1997; Goldstein *et al.* 1995; Rousset 1996; Slatkin 1995).

**Tabela 1** Programas de computador utilizados para analisar dados obtidos com microsatélites nucleares e citoplásticos

NOME DO PROGRAMA	CARACTERÍSTICAS	FONTE	REFERÊNCIA/CONTACTO
<b>MSA microsatellite analyser</b>	específico para genética de populações; IAM e SMM; grandes conjuntos de dados SSR; usa um formato simples do tipo Excel; apresenta as fórmulas	<a href="http://i12server.vu-wien.ac.at">http://i12server.vu-wien.ac.at</a>	(Dieringer & Schlötterer 2003).
<b>GeneAlex genetic analysis in Excel</b>	genética de populações; AMOVA; teste de Mantel; PCA; correlação espacial; determina a origem genética de indivíduos; formato simples no Excel; exporta para outros pacotes; limite de 126 loci SSR; não apresenta fórmulas; sem intervalos de confiança	<a href="http://www.anu.edu.au/BoZo/GenAlEx/">http://www.anu.edu.au/BoZo/GenAlEx/</a>	smouse@aesop.rutgers.edu.
<b>GENEPOP</b>	genética de populações; exporta para outros pacotes; efectua testes; apresenta as fórmulas	<a href="http://www.cefe.cnrs-mop.fr/">http://www.cefe.cnrs-mop.fr/</a>	(Raymond & Rousset 1995)
<b>RSTCALC</b>	estrutura da população, diferenciação genética e fluxo genético usando SSR e SMM; testa se os parâmetros são $\neq 0$ ; calcula intervalos de confiança; apresenta as fórmulas	<a href="http://helios.bto.ed.ac.uk/evolgen/rst/rst.html">http://helios.bto.ed.ac.uk/evolgen/rst/rst.html</a>	(Goodman 1997)
<b>STRUCTURE</b>	verifica se a população é estruturada; distribui indivíduos por populações; identifica migrantes; método “cluster” baseado num modelo baiano (probabilidades condicionadas); o ficheiro ‘input’ é trabalhoso	<a href="http://www.stats.ox.ac.uk/~pritch/home.html">http://www.stats.ox.ac.uk/~pritch/home.html</a>	(Pritchard <i>et al.</i> 2000).
<b>SPAGeDi</b>	estrutura genética espacial com diferentes tipos de marcadores; IAM e SMM; calcula intervalos de confiança; apresenta fórmulas. Max. 2000 loci e de 999 alelos por locus	<a href="http://www.ulb.ac.be/sciences/lagev/spagedi.html">http://www.ulb.ac.be/sciences/lagev/spagedi.html</a>	(Hardy & Vekemans 2002).
<b>MLTR Multilocus Mating System Program</b>	estima a taxa de fecundação cruzada através de informação multilocus; fornece erros padrões para as estimativas calculadas	<a href="http://genetics.forestry.ubc.ca/ritland/programs">http://genetics.forestry.ubc.ca/ritland/programs</a>	(Ritland 2002)
<b>GENELOSS</b>	simula os efeitos de “bottlenecks” na diversidade genética das populações com nuSSR	<a href="http://www.uow.edu.au/~pengland/docsetc.htm">http://www.uow.edu.au/~pengland/docsetc.htm</a>	pengland@uow.edu.au.
<b>BOTTLENECK</b>	detecta reduções no efectivo populacional a partir das frequências alélicas	<a href="http://www.ensam.inra.fr/URLB">http://www.ensam.inra.fr/URLB</a>	(Piry <i>et al.</i> 1999) (Cornuet & Luikart 1996)
<b>CERVUS</b>	Inferência da paternidade em populações naturais com dados codominantes; fácil de usar; 50 loci max; exporta para outros pacotes	<a href="http://helios.bto.ed.ac.uk/evolgen/cervus/cervusregister.html">http://helios.bto.ed.ac.uk/evolgen/cervus/cervusregister.html</a>	(Marshall <i>et al.</i> 1998).
<b>ZETA</b>	nº mínimo de nuSSR para identificar todos os indivíduos de uma amostra; cálculo do factor discriminante e do nº médio de alelos por SSR; Necessita de um compilador FORTRAN 95	Em construção!	sanchez@orleans.inra.fr.
<b>ARLEQUIN</b>	genética de populações; AMOVA; muito completo; vários tipos de marcadores e de parâmetros; apresenta as fórmulas; ficheiro ‘input’ complicado; dados haploides e diploides IAM and SMM	<a href="http://anthropologie.unige.ch/arlequin/">http://anthropologie.unige.ch/arlequin/</a>	(Michalakis & Excoffier 1996).
<b>FSTAT</b>	diversidade e diferenciação para marcadores codominantes e haplótipos; intervalos de confiança; testa parâmetros são $\neq 0$ ; 200 amostras, 200 loci e 999 alelos	<a href="http://www.unil.ch/izea/software/fstat.html">http://www.unil.ch/izea/software/fstat.html</a>	(Goudet 1995).
<b>HAPLODIV</b>	diversidade e diferenciação (cpSSR ou mtSSR); calcula os erros padrão; fórmulas no paper; 25 haplótipos e 200 populações	<a href="http://www.pierroton.inra.fr/genetics/labo/Software/">http://www.pierroton.inra.fr/genetics/labo/Software/</a>	(Pons & Petit 1995).
<b>TFPGA Tools for Population Genetic Analyses</b>	genética de populações; UPMGA; teste de Mantel; formato simples; não apresenta fórmulas	<a href="http://bioweb.usu.edu/mpmbio/">http://bioweb.usu.edu/mpmbio/</a>	Mark.Miller@cnr.usu.edu.
<b>FAMAZ</b>	usa o método de “likelihood” e simulação para análise de parentesco com dados codominantes, dominantes, haploides e mistos; complicado de instalar	<a href="http://www.pierroton.inra.fr/genetics/labo/Software/">http://www.pierroton.inra.fr/genetics/labo/Software/</a>	(Gerber <i>et al.</i> 2003).

## Aplicações

A nível do indivíduo, os microsatélites nucleares são utilizados para identificação individual, variação somaclonal, identificação de germoplasma e análise parental (“fingerprinting”, Rajora & Rahman (2003); variação somaclonal Rahman & Rajora (2001); identificação de germoplasma, Khadari *et al.* (2003); análise de paternidade, Gerber *et al.* (2000)). Estes marcadores têm, também, sido usados em

estudos ao nível das populações e ecologia: estrutura genética, fluxo genético, efectivo populacional e migração (ex: Bagley *et al.* (1999); Dow & Ashley (1996); Gibbs & Weatherhead (1999); Ribeiro *et al.* (2001)). Podemos ainda referir o seu uso como pontos de âncora codominantes em mapeamento genético e sintetia (ex: Marques *et al.* (2002)), em sistemas de cruzamento e em estudos de contaminação de pólen (Collevatti *et al.* 2003; Plomion *et al.* 2001), nos recursos genéticos e na identificação de proveniências (nuSSR, Derory *et al.* (2002); cpSSR, Ribeiro *et al.* (2002). González-Martínez *et al.* (submetido), não encontraram nenhuma correlação quando compararam a distribuição da diversidade entre vários marcadores moleculares, incluindo microsátélites, e características morfológicas adaptativas em pinheiro bravo. No entanto, na mesma espécie, Ribeiro *et al.* (submetido) verificaram a existência de uma correlação positiva entre marcadores moleculares (cpSSR) e quantitativos, usando a diversidade genética intra populacional, o que parece ser devido aos elevados níveis de fluxo genético encontrados numa escala local. Novas estratégias são necessárias para identificar e analisar microsátélites ligados a regiões do genoma relevantes do ponto de vista adaptativo e, para isso, um uso combinado de SSRs e SNPs parece ser promissor.

## Análise de dados obtidos com SSR

A análise dos dados vem na sequência da(s) pergunta(s) que formulámos quando iniciámos o estudo. Será que queríamos estudar a estrutura de uma espécie ou efectuar análises de paternidade? É importante que na base dos dados obtidos esteja um bom delineamento experimental: o número de populações, o número de indivíduos por população, ou outros aspectos pertinentes para o estudo. A análise que iremos efectuar é função da(s) pergunta(s) e dos parâmetros a estimar: heterozigocidade, sistemas de cruzamento, estrutura da população, diferenciação, filogenia, “fingerprinting”, ou outros. Quando iniciamos um estudo deveremos ter em conta os seguintes aspectos: i) obter uma amostra representativa (da população e do genoma), ii) utilizar marcadores adequados, iii) obter um conjunto de dados isentos de erros, iv) estabelecer os parâmetros a determinar, v) estimar os parâmetros (modelo de mutação adequado), vi) obter medidas de confiança das estimativas calculadas e, vii) escolher um programa (ou diferentes programas) para analisar os dados (alguns exemplos de pacotes estatísticos encontram-se na tabela 1). Para obter mais informações sobre estes e outros pacotes estatísticos apropriados para a análise de dados obtidos com microsátélites consultar o artigo de Luikart & England (1999).

## Estudo-exemplo

No artigo de England *et al.* (2002) encontramos um exemplo do uso de microsátélites para o estudo da espécie rara australiana *Grevillea macleayana*. Nesse estudo, os autores queriam saber qual foi a influência da distribuição fragmentária desta espécie na estrutura genética e se poderia ter causado uma depleção na variabilidade genética. Escolheram plantas adultas, recolhidas nas áreas Norte, Centro e Sul da distribuição da espécie. Os objectivos do estudo foram: (i) quantificar os níveis de variação genética entre e dentro das populações de cada região amostrada; (ii) examinar os padrões de estrutura genética; e (iii) inferir a importância do fluxo genético do passado. Para análise dos dados os autores usaram o programa **GENEPOP** para calcular parâmetros de diversidade genética, testar a deficiência em heterozigóticos, o desequilíbrio de ligamento e a heterogeneidade génica e genotípica entre populações. O  $\rho_{ST}$  foi calculado no programa **RSTCALC**. Este programa foi também usado para calcular o parâmetro  $N_m$  (número de migrantes por geração) a partir do  $\rho_{ST}$ . Segundo os autores, embora o  $\rho_{ST}$ , um estimador centrado do  $R_{ST}$  (Slatkin 1995) e análogo do  $F_{ST}$  (Wright 1965) seja apropriado para *loci* em mutação rápida, como é o caso dos microsátélites, o  $F_{ST}$  também foi calculado, porque é muito usado noutros estudos e porque os *loci* estudados eram relativamente invariantes. Usaram o **TFGA** para calcular a distância genética de Nei (Nei 1978) e para calcular um fenograma de grupos UPMGA baseado em 1000 árvores permutadas fazendo “bootstrap” usando os *loci*. A opção Mantel do **GENEPOP** foi usada para testar o isolamento por distância do  $F_{ST}$  e do  $\rho_{ST}$ . Usaram o programa **BOTTLENECK** para determinar se as populações teriam sofrido recentemente uma “bottleneck” e se estariam, por isso, fora do equilíbrio mutação-deriva. Com os seis SSR utilizados, os autores encontraram uma diversidade relativamente baixa dentro das populações, uma diferenciação populacional significativa e uma estrutura genética moderada, demonstrando a existência de

isolamento por distância consistente com um baixo fluxo genético. A distribuição das frequências dos tamanhos dos alelos sugere que a diferenciação geográfica foi causada por mutação. Foi observada um desvio ao equilíbrio mutação-deriva nalgumas populações, o que sugere que estas sofreram o efeito de “bottlenecks”. Os padrões naturais de dispersão limitada do pólen e semente, juntamente com a distribuição fragmentada e modelada pelo fogo desta espécie, podem ter limitado o fluxo genético entre as populações, no passado.

## Considerações finais

Existe um crescimento explosivo de novas possibilidades de análise de dados devido ao aumento da quantidade e qualidade dos dados (disponibilidade de mais *loci*, mais polimórficos e mais abundantes ao longo do genoma e maiores possibilidades de automatização), a pacotes estatísticos disponíveis mais inovativos e poderosos (metodologias mais eficientes para simulações estocásticas – ex. Cadeias de Markov - Monte Carlo), a computadores mais rápidos, ao acesso fácil e gratuito através da Web e ao contacto fácil com os autores dos programas. Como afirmam Luikart & England (1999): “It is indeed an exciting time to be a population geneticist”.

## Referências

- Bachtrog D., Agis M., Imhof M., Schlötterer C. (2000) Microsatellite variability differs between dinucleotide repeat motifs-evidence from *Drosophila melanogaster*. *Molecular Biology and Evolution* **17**, 1277-1285.
- Bagley M.J., Lindquist D.G., Geller J.B. (1999) Microsatellite variation, effective population size, and population genetic structure of vermilion snapper, *Rhomboplites aurorubens*, of the southeastern USA. *Marine Biology* **134**, 609-620.
- Balloux F., Brünnner H., Lugo-Moulin N., Hausser J., Goudet J. (2000) Microsatellites can be misleading: an empirical and simulation study. *Evolution* **54**, 1414-1422.
- Balloux F., Lugo-Moulin N. (2002) The estimation of population differentiation with microsatellite markers. *Molecular Ecology* **11**, 155-165.
- Collevatti R.G., Grattapaglia D., Duvall J. (2003) High resolution microsatellite based analysis of the mating system allows the detection of significant biparental inbreeding in *Caryocar brasiliense*, an endangered tropical tree species. *Heredity* **86**, 60-67.
- Cornuet, J.M., Luikart G. (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* **144**, 2001-2014.
- Crow J.F., Kimura M. (1970) *An introduction to population genetics theory* Harper & Row, New York.
- Derory J., Mariette S., González-Martínez S.C., Chagné D., Madur D., Gerber S., Ribeiro M.M., Plomion C. (2002) What can nuclear microsatellites tell us about maritime pine genetic resources conservation and provenances certification strategies? *Annals of Forest Science* **59**, 699-708.
- Dieringer D., Schlötterer C. (2003) Microsatellite analyser (MSA): a platform independent analysis tool for large microsatellite data sets. *Molecular Ecology Notes* **3**, 167-169.
- Dow B., Ashley M. (1996) Microsatellite analysis of seed dispersal and parentage of saplings in bur oak, *Quercus macrocarpa*. *Molecular Ecology* **5**, 615-627.
- Ellegren H. (2000) Heterogenous mutation processes in human microsatellite DNA sequences. *Nature Genetics* **24**, 400-402.
- England P.R., Usher A.V., Whelan R.J., Ayre D.J. (2002) Microsatellite diversity and genetic structure of fragmented populations of the rare, fire-dependent shrub *Grevillea macleayana*. *Molecular Ecology* **11**, 967-977.
- Estoup A., Jarne J., Cournet J.M. (1995) Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Molecular Ecology* **11**, 1591-1604.
- Feldman M.W., Bergman A., Pollock D.D., Goldstein D.B. (1997) Microsatellite genetic distances with range constraints: Analytic description and problems of estimation. *Genetics* **145**, 207-216.

- Garza J.C., Slatkin M., Freimer N.B. (1995) Microsatellite allele frequencies in humans and chimpanzees, with implications for constraints on allele size. *Molecular Biology and Evolution* **12**, 594-603.
- Gerber S., Chabrier P., Kremer A. (2003) FaMoz: a software for parentage analysis using dominant, codominant and uniparentally inherited markers. *Molecular Ecology Notes* **in press**.
- Gerber S., Mariette S., Streiff R., Bodénès C., Kremer A. (2000) Comparison of microsatellites and amplified fragment length polymorphism markers for parentage analysis. *Molecular Ecology* **9**, 1037–1048.
- Gibbs H.L., Weatherhead P.J. (1999) Insight into population ecology and sexual selection in snakes through the application of DNA-based markers. *The Journal of Heredity* **92**, 173-179.
- Goldstein D.B., Linares A.R., Cavalli-Sforza L.L., Feldman M.W. (1995) An evaluation of genetic distances for use with microsatellite loci. *Genetics* **139**, 463-471.
- González-Martínez S.C., Mariette S., Ribeiro M.M., Burbán C., Raffin A., Chambel M.R., Ribeiro C., Aguiar A., Plomion C., Alía R., Gil L., Vendramin G.G., Kremer A. (submitted) Genetic resources in maritime pine (*Pinus pinaster* Aiton): patterns of differentiation and correlation between molecular and quantitative measures of genetic variation. *Forest Ecology and Management*.
- Goodman S.J. (1997) Rst Calc: A collection of computer programs for calculating unbiased estimates of genetic differentiation and determining their significance for microsatellite data. *Molecular Ecology* **6**, 881-885.
- Goudet J. (1995) FSTAT (Version 1.2): A computer program to calculate F-statistics. *Journal of Heredity* **86**, 485-486.
- Hardy O.J., Vekemans X. (2002) Spagedi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes* **2**, 618 -620.
- Hedrick R.P. (1999) Perspective: highly variable loci and their interpretation in evolution and conservation. *Evolution* **53**, 313-318.
- Karp A., Edwards K.J. (1997) DNA markers: a global overview. In: *DNA markers: protocols, applications and overviews* eds. Caetano-Anollés G., Gresshoff P.M.), pp. 1-13. Wiley-VCH, New York.
- Khadari B., Breton C., Moutier N., Roger J.P., Besnard G., Bervillé A., Dosba F. (2003) The use of molecular markers for germplasm management in a French olive collection. *Theoretical and Applied Genetics* **106**, 521–529.
- Kimura M., Crow J.F. (1964) The number of alleles that can be maintained in a finite population. *Genetics* **49**, 725-738.
- Kimura M., Ohta T. (1978) Stepwise mutation model and distribution of allelic frequencies in a finite population. *Proceedings of the National Academy of Sciences of USA* **75**, 2868-2872.
- Liepelt S., Kuhlenskamp V., Anzidei M., Vendramin G.G., Ziegenhagen B. (2001) Pitfalls in determining size homoplasy of microsatellite loci. *Molecular Ecology Notes* **1**, 332-335.
- Luikart G., England P.R. (1999) Statistical analysis of microsatellite DNA data. *Trends in Ecology & Evolution* **14**, 253-256.
- Marques C.M., Carocha V.J., Brondani R.P.V., Grattapaglia D., Sederoff R. (2002) Conservation and synteny of SSR loci and QTL for vegetative propagation in four *Eucalyptus* species. *Theoretical and Applied Genetics* **105**, 474-478.
- Marshall T.C., Slate J., Kruuk L., Pemberton J.M. (1998) Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular Ecology* **7**, 639-655.
- Michalakis Y., Excoffier L. (1996) A generic estimation of population subdivision using distances between alleles with special reference for microsatellite loci. *Genetics* **142**, 1061-1064.
- Mitton J.B. (1994) Molecular approaches to population biology. *Annual Review of Ecology and Systematics* **25**, 45-69.
- Nei M. (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* **89**, 583-590.
- Parker P.G., Snow A.A., Schug M.D., Booton G.C., Fuerst P.A. (1998) What molecules can tell us about populations: choosing and using a molecular marker. *Ecology* **79**, 361-382.
- Petes T.D., Greewell P.W., Dominska M. (1997) Stabilization of microsatellite sequences by variant repeats in the yeast *Saccharomyces cerevisiae*. *Genetics* **146**, 491-498.

- Piry S., Luikart G., Cornuet J.M. (1999) BOTTLENECK: A computer program for detecting recent reductions in the effective population size using allele frequency data. *Journal of Heredity* **90**, 502-503.
- Plomion C., LeProvost G., Pot D., Vendramin G., Gerber S., Decroocq S., Brach J., Raffin A., Pastuszka P. (2001) Pollen contamination in a maritime pine polycross seed orchard and certification of improved seeds using chloroplast microsatellites. *Canadian Journal of Forest Research* **31**, 1816-1825.
- Pons O., Petit R.J. (1995) Estimation, variance and optimal sampling of gene diversity .1. Haploid locus. *Theoretical and Applied Genetics* **90**, 462-470.
- Primmer C.R., Saino N., Moller A.P., Ellegren H. (1998) Unraveling the processes of microsatellite evolution through analysis of germ line mutations in barn swallows *Hirundo rustica*. *Molecular Biology and Evolution* **15**, 1047-1054.
- Pritchard J.K., Stephens M., Donnelly P. (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**, 945-959.
- Rahman M.H., Rajora O.P. (2001) Microsatellite DNA somaclonal variation in micropropagated trembling aspen (*Populus tremuloides*). *Plant Cell Reports* **20**, 531-536.
- Rajora O.P., Rahman M.H. (2003) Microsatellite DNA and RAPD fingerprinting, identification and genetic relationships of hybrid poplar (*Populus x canadensis*) cultivars. *Theoretical and Applied Genetics* **106**, 470-477.
- Raymond M., Rousset F. (1995) GENEPOP (version 1.2): Population genetics software for exact tests and ecumenicism. *Journal of Heredity* **86**, 248-249.
- Ribeiro M.M., González-Martínez S.C., Aguiar A., Plomion C., Kremer A., Alia R. (submitted) Genetic variation in quantitative traits and cpSSR loci in Portuguese maritime pine populations. *Heredity*.
- Ribeiro M.M., LeProvost G., Gerber S., Vendramin G.G., Anzidei M., Decroocq S., Marpeau A., Mariette S., Plomion C. (2002) Origin identification of maritime pine stands in France using chloroplast simple-sequence repeats. *Annals of Forest Science* **59**, 53-62.
- Ribeiro M.M., Plomion C., Petit R., Vendramin G.G., Szmidt A.E. (2001) Variation of chloroplast simple-sequence repeats in Portuguese maritime pine (*Pinus pinaster* Ait.). *Theoretical and Applied Genetics* **102**, 97-103.
- Ritland K. (2002) Extensions of models for the estimation of mating systems using n independent loci. *Heredity* **88**, 221-228.
- Rousset F. (1996) Equilibrium values of measures of population subdivision for stepwise mutation processes. *Genetics* **142**, 1357-1362.
- Schlötterer C., Ritter R., Harr B., Brem G. (1998) High mutation rate of a long microsatellite allele in *Drosophila melanogaster* provides evidence for allele-specific mutation rates. *Molecular Biology and Evolution* **15**, 1269-1274.
- Shriver M.D., Jin L., Boerwinkle L.E., Deka R., Ferrel R.E., Chakraborty R. (1995) A novel measure of genetic distance for highly polymorphic tandem repeat loci. *Molecular Biology and Evolution* **12**, 914-920.
- Slatkin M. (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics* **139**, 457-462.
- Strand M., Prolla T.A., Liskay R.M., Petes T.D. (1993) Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature* **365**, 274-276.
- Szmidt A.E., Wang X.-R. (2000) Genetic markers in forest genetics and breeding - the tunnel remains dark. In: *Forest Genetics and Sustainability* (ed. Matyas C.), pp. 31-48. Kluwer Academic Publishers, The Netherlands.
- Taberlet P., Luikart G. (1999) Non-invasive genetic sampling and individual identification. *Biological Journal of the Linnean Society* **68**, 41-55.
- Tautz D. (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acid Research* **17**, 6463-6471.
- Vekemans X., Jacquemart A.L. (1997) Perspectives on the use of molecular markers in plant population biology. *Belgian Journal of Botany* **129**, 91-100.
- Wang X.-R., Szmidt A.E. (2001) Molecular markers in population genetics of forest trees. *Scandinavian Journal of Forest Research* **16**, 199 - 220.



Wright S. (1965) The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution* **19**, 395-420.